CONFIDENCE BOUNDS ON CANONICAL REGRESSIONS[1]

by

S. N. ROY

INSTITUTE OF STATISTICS, UNIVERSITY OF NORTH CAROLINA

Institute of Statistics
Mimeograph Series No. 98

March 11, 1954

[1] Work sponsored by the Office of Naval Research under Contract No. N7-onr-28402.

CONFIDENCE BOUNDS ON CANONICAL REGRESSIONS[1]

BY

S. N. ROY
Institute of Statistics, University of North Carolina

1. Summary. This paper starts from a (p+q)-variate normal population ($p \leq q$) with a p.d. dispersion matrix consisting of submatrices $\Sigma_{11}(p \times p)$, $\Sigma_{22}(q \times q)$, $\Sigma_{12}(p \times q)$ which stand respectively for the dispersion matrix of the p-set, the q-set and that between the p-set and the q-set, and then defines, in a natural manner, the matrix of regression of the p-set on the q-set, in the form $\Sigma_{12}\Sigma_{22}^{-1}$. This matrix is denoted by $\beta(p \times q)$ and a bilinear function $\underline{d}_1'(1 \times p)\beta(p \times q)\underline{d}_2(q \times 1)$ is considered where $\underline{d}_1(p \times 1)$ and $\underline{d}_2(q \times 1)$ are two arbitrary vectors, each of unit modulus. Simultaneous confidence bounds are given on all such bilinear compounds $\underline{d}_1'\beta\underline{d}_2$ with a joint confidence coefficient greater than or equal to a preassigned level. For this purpose certain results and techniques are used which were discussed in previous papers [1,2,3.]

2. Introduction. We recall the confidence statement [1, (4.3.4)], with a confidence coefficient $1 - \alpha$:

$$(2.1) \quad b - \frac{t_\alpha(n-2)}{\sqrt{n-2}}(1 - r^2)^{\frac{1}{2}} \frac{s_1}{s_2} \leq \beta \leq b + \frac{t_\alpha(n-2)}{\sqrt{n-2}}(1 - r^2)^{\frac{1}{2}} \frac{s_1}{s_2},$$

where β (which is now a scalar) stands for the population regression of x_1 on x_2 (where x_1 and x_2 have a bivariate normal distribution), b for the sample regression (in a random sample of size $n \geq 3$), r

---

[1] Work sponsored by the Office of Naval Research under Contract No. N7-onr-28402.

for the sample correlation, s_1 and s_2 for the two sample standard deviations and t_α for the upper $\alpha/2$-point of the t-distribution with D.F.(n-2).

We also note that

(2.2) $$b = rs_1/s_2 = rs_1s_2/s_2^2 \text{ and } \beta = \rho\sigma_1\sigma_2/\sigma_2^2 ,$$

where ρ, σ_1, and σ_2 stand respectively for the population correlation coefficient and the two standard deviations.

Denoting by C(M) the characteristic root of a p x p matrix (whose elements are real or complex numbers) we recall also [2, (1.2)] that, if A(p x q) and B(q x p) are two such matrices, then

(2.3) $$C(AB) = C(BA),$$

meaning thereby that any non-zero root of (AB) is also a non-zero root of (BA) and vice versa.

We also note that

(2.4) $$\mathrm{tr}\ (AB) = \mathrm{tr}\ (BA).$$

We further recall [2, (2.2.4)] that if A and B are two p x p hermitian matrices, one of which, say A, is p.d. and the other, i.e., B, at least p.s.d., then, denoting by c_{max} and c_{min} the largest and the smallest characteristic roots, we have

(2.5) $$c_{min}(A)\ c_{min}(B) \leq \text{all } c\ (AB) \leq c_{max}(A)\ c_{max}(B).$$

We next recall [3, first paragraph of section 3] that

(2.6) $$\text{"If } E_1, \text{ then } E_2\text{"} \Longrightarrow \text{" } E_1 \subset E_2\text{"} \longrightarrow P(E_1) \leq P(E_2).$$

We now start [1, section 6.2] with a random sample of size n (> p+q; $p \leq q$) from a (p+q)-variate normal population, and next reduce for the means and set

$$(n-1)\begin{array}{cc} & \begin{array}{cc} p & q \end{array} \\ \begin{array}{c} p \\ q \end{array} & \begin{pmatrix} S_{11} & S_{12} \\ S_{12}' & S_{22} \end{pmatrix} \end{array} = \begin{array}{cc} & n-1 \\ \begin{array}{c} p \\ q \end{array} & \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \end{array} \begin{array}{cc} \begin{pmatrix} Y_1' & Y_2' \end{pmatrix} & n-1 \\ \begin{array}{cc} p & q \end{array} & \end{array},$$

where S_{11}, S_{22} and S_{12} stand respectively for the sample dispersion sub-matrices of the p-set, the q-set and that between the p-set and the q-set and where Y_1 and Y_2 have the p.d.f.

$$(2.7) \qquad \text{const.}\ \exp\left[-\tfrac{1}{2}\,\mathrm{tr}\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix}^{-1}\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}\begin{pmatrix} Y_1' & Y_2' \end{pmatrix}\right].$$

We next recall [1, section 6.2] that there exist non-singular $\mu_1(p \times p)$ and $\mu_2(q \times q)$ such that

$$(2.8) \quad \Sigma_{11}(p \times p) = \mu_1(p \times p)\,\mu_1'(p \times p),\ \Sigma_{22}(q \times q) = \mu_2(q \times q)\,\mu_2'(q \times q)$$

$$\text{and } \Sigma_{12}(p \times q) = \mu_1(p \times p)\,\begin{array}{cc}(D_{\sqrt{\theta}} & 0) \\ p & q-p\end{array}\,\mu_2'(q \times q),$$

where $D_{\sqrt{\theta}}$ stands for a diagonal matrix the squares of whose diagonal elements are the (all non-negative) characteristic roots of the matrix $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}'$ (i.e., the squares of the population canonical correlations between the p-set and the q-set). As in [1, section 6.2], denoting by I(m) an m x m identity matrix, we have

$$(2.9)\quad \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix}^{-1} = \left[\begin{pmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{pmatrix} \left(\begin{array}{c|c} I(p) & \begin{pmatrix} D_{\sqrt{\Theta}} & 0 \end{pmatrix} \\ \hline \begin{pmatrix} D_{\sqrt{\Theta}} \\ 0 \end{pmatrix} & I(q) \end{array} \right) \begin{pmatrix} \mu_1' & 0 \\ 0 & \mu_2' \end{pmatrix} \right]^{-1}$$

$$= \left[\begin{pmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{pmatrix} \left(\begin{array}{c|c} D_{\sqrt{1-\Theta}} & \begin{pmatrix} D_{\sqrt{\Theta}} & 0 \end{pmatrix} \\ \hline 0 & I(q) \end{array} \right) \left(\begin{array}{c|c} D_{\sqrt{1-\Theta}} & 0 \\ \hline \begin{pmatrix} D_{\sqrt{\Theta}} \\ 0 \end{pmatrix} & I(q) \end{array} \right) \begin{pmatrix} \mu_1' & 0 \\ 0 & \mu_2' \end{pmatrix} \right]^{-1}$$

$$= \begin{pmatrix} \mu_1'^{-1} & 0 \\ 0 & \mu_2'^{-1} \end{pmatrix} \left(\begin{array}{c|c} D_{\sqrt{1/1-\Theta}} & 0 \\ \hline -\begin{pmatrix} D_{\sqrt{\Theta/1-\Theta}} \\ 0 \end{pmatrix} & I(q) \end{array} \right) \left(\begin{array}{c|c} D_{\sqrt{1/1-\Theta}} & -\begin{pmatrix} D_{\sqrt{\Theta/1-\Theta}} & 0 \end{pmatrix} \\ \hline 0 & I(q) \end{array} \right)$$

$$\times \begin{pmatrix} \mu_1^{-1} & 0 \\ 0 & \mu_2^{-1} \end{pmatrix} .$$

Going back to (2.7) and using (2.4) we have now

$$(2.10)\quad \operatorname{tr} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix}^{-1} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \begin{pmatrix} Y_1' & Y_2' \end{pmatrix}$$

$$= \operatorname{tr} \left(\begin{array}{c|c} D_{\sqrt{1/1-\Theta}} & -\begin{pmatrix} D_{\sqrt{\Theta/1-\Theta}} & 0 \end{pmatrix} \\ \hline 0 & I(q) \end{array} \right) \begin{pmatrix} \mu_1^{-1} & 0 \\ 0 & \mu_2^{-1} \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \begin{pmatrix} Y_1' & Y_2' \end{pmatrix}$$

$$\times \begin{pmatrix} \mu_1'^{-1} & 0 \\ 0 & \mu_2'^{-1} \end{pmatrix} \left(\begin{array}{c|c} D_{\sqrt{1/1-\Theta}} & 0 \\ \hline -\begin{pmatrix} D_{\sqrt{\Theta/1-\Theta}} \\ 0 \end{pmatrix} & I(q) \end{array} \right) = \operatorname{tr} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \begin{pmatrix} Z_1' & Z_2' \end{pmatrix} ,$$

where

$$(2.11) \qquad Z_1 = D_{\sqrt{1/1-\Theta}}\ \mu_1^{-1}\ Y_1 - \left(D_{\sqrt{\Theta/1-\Theta}} \quad 0\right) \mu_2^{-1}\ Y_2\ ,$$

$$Z_2 = \mu_2^{-1}\ Y_2\ .$$

Thus it is easy to check from (2.7), (2.10) and (2.11) that (Z_1, Z_2) have the p.d.f.

$$(2.12) \qquad \text{const. exp.} \ -\tfrac{1}{2}\ \text{tr} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} (Z_1' \quad Z_2')\ .$$

Consider now, for any two arbitrary non-null vectors $\underline{a}_1(p \times 1)$ and $\underline{a}_2(q \times 1)$ and for a fixed positive Θ_0, the statement

$$(2.13) \qquad \frac{(\underline{a}_1'\ Z_1\ Z_2'\ \underline{a}_2)^2}{(\underline{a}_1'\ Z_1\ Z_1'\ \underline{a}_1)(\underline{a}_2'\ Z_2\ Z_2'\ \underline{a}_2)} \leq \Theta_0\ ,$$

which can be written in terms of Y_1 and Y_2 as

$$(2.14) \qquad \frac{\left[\underline{a}_1'\ D_{\sqrt{1/1-\Theta}}\ \mu_1^{-1} Y_1 Y_2' \mu_2'^{-1} - \left(D_{\sqrt{\Theta/1-\Theta}} \quad 0\right) \mu_2^{-1} Y_2 Y_2' \mu_2'^{-1}\ \underline{a}_2\right]^2}{(\underline{a}_2' \mu_2^{-1} Y_2 Y_2' \mu_2'^{-1} \underline{a}_2)(\underline{a}_1' Q Q' \underline{a}_1)} \leq \Theta_0\ ,$$

where

$$(2.15) \qquad Q = D_{\sqrt{1/1-\Theta}}\ \mu_1^{-1} Y_1 - \left(D_{\sqrt{\Theta/1-\Theta}} \quad 0\right) \mu_2^{-1} Y_2\ .$$

Now putting

$$(2.16)\quad \underline{b}_1'(1 \times p) = \underline{a}_1' D_{\sqrt{\Theta/1-\Theta}}\, \mu_1^{-1} \text{ and } \underline{b}_2'(1 \times q) = \underline{a}_2'\mu_2^{-1},$$

and using (2.8) and (2.6), we check that (2.14) reduces to

$$\frac{[\underline{b}_1'(Y_1Y_2' - \beta Y_2Y_2')\underline{b}_2]^2}{(\underline{b}_2'Y_2Y_2'\underline{b}_2)[\underline{b}_1'(Y_1-\beta Y_2)(Y_1' - Y_2'\beta')\underline{b}_2]} \leq \Theta_0$$

or

$$\frac{[\underline{b}_1'(S_{12} - \beta S_{22}')\underline{b}_2]^2}{(\underline{b}_2'S_{22}\underline{b}_2)[\underline{b}_1'(S_{11}-S_{12}\beta' - \beta S_{12}'+\beta S_{22}\beta')\underline{b}_1]} \leq \Theta_0,$$

where

$$(2.17)\quad \beta(p \times q) = \mu_1\begin{pmatrix} D_{\sqrt{\Theta}} & 0 \end{pmatrix}\mu_2^{-1} = \Sigma_{12}\Sigma_{22}^{-1}.$$

β defined by (2.17) can be appropriately called the matrix of population regression of the p-set on the q-set and it is the only set of population parameters that occurs in the statement (2.17).

3. <u>Confidence bounds on the regression matrix β.</u>

It is well known [1] that the statement (2.17), for all arbitrary non-null $\underline{b}_1$ and $\underline{b}_2$, is exactly equivalent to

(3.1) all Θ_i's $\leq \Theta_0$ or $\Theta_p \leq \Theta_0$,

where Θ_i's $(i = 1, 2, \ldots, p;\ 0 \leq \Theta_1 \leq \ldots \leq \Theta_p \leq 1)$ are the roots of the determinantal equation in Θ:

$$(3.2)\quad \left|\Theta(S_{11}-S_{12}\beta'-\beta S_{12}'+\beta S_{22}\beta') - (S_{12}-\beta S_{22})S_{22}^{-1}(S_{12}'-S_{22}\beta')\right| = 0.$$

Now put $\lambda = \theta/1 - \theta$, so that we have from (3.2), the determinantal equation in λ

$$(3.3)\quad \left| \lambda(S_{11}-S_{12}S_{22}^{-1}S_{12}') - (S_{12}S_{22}^{-1} - \beta)S_{22}(S_{22}^{-1}S_{12}' - \beta') \right| = 0.$$

The statement (3.1) can now be replaced by the statement that

(3.4) the largest characteristic root $\leq \theta_0/1 - \theta_0$,

i.e.,

$$(3.5)\quad \text{all } c\,[(S_{11}-S_{12}S_{22}^{-1}S_{12}')^{-1}(B-\beta)S_{22}(B'-\beta')] \leq \theta_0/(1-\theta_0),$$

where

$$(3.6)\quad B(p \times q) = S_{12}S_{22}^{-1},$$

which may be appropriately called the matrix of sample regression of the p-set on the q-set.

We note that (3.5) $\Longrightarrow$ (3.1) $\Longrightarrow$ (2.13), so that θ_p is the largest characteristic root of the matrix $(Z_1Z_1')^{-1}(Z_1Z_2')(Z_2Z_2')^{-1}(Z_2Z_1')$, where (Z_1, Z_2) have the p.d.f. (2.12). The joint distribution of these central θ_i's, and also of the largest root θ_p being known, all that we have to do to make (3.5), i.e., (3.1), i.e., (2.13), a simultaneous confidence statement with a joint confidence coefficient $1 - \alpha$ is to choose $\theta_0 = \theta_\alpha(p,q,n-1)$ where the quantity on the right hand side is defined by

$$(3.7)\quad P\,(\text{central } \theta_p \geq \theta_0) = \alpha.$$

Substituting now $\theta_\alpha(p, q, n-1)$ (to be sometimes denoted more simply by θ_α) for θ_0 in (3.5), we have a simultaneous confidence statement with a joint confidence coefficient $1 - \alpha$:

Now applying (2.3), (2.5) and (2.6) (in the same manner as in [3]), we have from (3.5), now with a joint confidence coefficient $\geq 1 - \alpha$, the following simultaneous confidence statement

$$(3.8) \quad \text{all } c\,[(B - \beta)(B' - \beta')] \leq \frac{\theta_\alpha}{1-\theta_\alpha}\; c_{max}(S_{11}-S_{12}S_{22}^{-1}S_{12}') \times c_{max}(S_{22}^{-1}).$$

Now note that $c_{max}(S_{22}^{-1}) = 1/c_{min}(S_{22})$,

$c_{max}(S_{11}-S_{12}S_{22}^{-1}S_{12}') \leq c_{max}(S_{11})\; c_{max}(I - S_{11}^{-1}S_{12}S_{22}^{-1}S_{12}')$ and

$c_{max}(I - S_{11}^{-1}S_{12}S_{22}^{-1}S_{12}') = 1 - c_{min}(S_{11}^{-1}S_{12}S_{22}^{-1}S_{12}')$.

Using these, we check that (3.8) can be replaced by the following (with a confidence coefficient $\geq 1 - \alpha$):

$$(3.9) \quad \text{all } c\,[(B - \beta)(B' - \beta')] \leq \frac{\theta_\alpha}{1-\theta_\alpha}\; [1 - c_{min}(S_{11}^{-1}S_{12}S_{22}^{-1}S_{12}')]$$

$$\times\; c_{max}(S_{11})/c_{min}(S_{22}).$$

We next recall the following two well-known results (repeatedly used in [1]):

(3.10) all $c(M) \leq g$ (for a $p \times p$ real matrix M with real roots)

$\Longrightarrow$ $\underline{d}_1'\,(1 \times p)\; M(p \times p)\; \underline{d}_1(p \times 1)$ (for all arbitrary unit vectors $\underline{d}_1$) $\leq g$ and

(3.11) $\underline{x}'\,(1 \times q)\; \underline{x}(q \times 1) \leq h(\, > 1) \Longrightarrow |\underline{x}_2'(1 \times q)\; \underline{d}_2(q \times 1)| \leq \sqrt{h}$

(for all arbitrary unit vectors $\underline{d}_2$).

Applying (3.10) and (3.11) to (3.9) we have (with a joint confidence coefficient $\geq 1 - \alpha$) the following simultaneous confidence statement (for

all arbitrary unit vectors $\underline{d}_1(p \times 1)$ and $\underline{d}_2(q \times 1)$),

(3.12) $\left| \underline{d}_1' (B - \beta) \underline{d}_2 \right| \leq$ [right hand side of (3.9)]$^{\frac{1}{2}}$,

or ultimately

(3.13) $\underline{d}_1' B\underline{d}_2 - \sqrt{E} \leq \underline{d}_1' \beta\underline{d}_2 \leq \underline{d}_1 B\underline{d}_2 + \sqrt{E}$,

where

(3.14) $E = [Q_\alpha/(1 - Q_\alpha)] [1 - c_{min}(S_{11}^{-1}S_{12}S_{22}^{-1}S_{12}')] [c_{max}(S_{11})/c_{min}(S_{22})]$.

A set of simultaneous confidence bounds on just the elements β_{ij} of the β-matrix would be a subset of the bounds on the total set $\underline{d}_1'\beta\underline{d}_2$. It is worthwhile to check that if $p = q = 1$, (3.13) reduces, as it should, to (2.1). Also if $p = 1$, we should have another special case of (3.13) giving a set of simultaneous confidence bounds on all linear functions of the partial regressions of one variate on several others. Thus, in several ways, (3.13) seems to be an appropriate generalization of (2.1).

References

1. S. N. Roy and R. C. Bose, "Simultaneous confidence interval estimation," Annals of Mathematical Statistics, Vol. 24 (1953), pp. 513-36.

2. S. N. Roy, "A useful theorem in matrix algebra," mimeographed paper.

3. S. N. Roy, "Some further results in simultaneous confidence interval estimation," mimeographed paper.